WHAT IS CLAIMED:

1.   A method of creating a product catalog stored on computer readable media by aggregating product information from a plurality of product information sources having disparate formats for product information and storing the information in a taxonomy, said method comprising:

processing plural product information records from the product information sources into one or more groups based on which product information records are likely to correspond to the same product;

correlating a unique product ID corresponding to the product associated with each of said groups to identify the product;

comparing each identified product to categories of a taxonomy to determine a category for the identified products in the taxonomy; and

determining attributes for each categorized product based on the product information records corresponding to each group;

creating product specifications based on the determined attributes; and

storing the product specification in the corresponding determined categories of the taxonomy.

2.   A method as recited in claim 1, wherein said processing step includes determining which products referred to in said product information records are likely to be the same by comparing data strings associated with the products, and determining a common data string.

3.   A method as recited in claim 2, wherein said data strings include at least one of manufacturer part numbers, model identifiers and uniform product codes.

4.   A method as recited in claim 1, wherein said processing step comprises comparing product names in the product information records and grouping together all products having a substantially similar product name.

5.   A method as recited in claim 4, wherein said product names are compared without regard for differences in capitalization and punctuation.

NVA276042.1

6.   A method as recited in claim 4, further comprising a second processing step wherein generic nouns associated with the products are parsed and ignored in determining the groups.

7.   A method as recited in claim 1, wherein said processing step comprises parsing selected adjectives and ignoring the selected adjectives in determining the groups.

8.   A method as recited in claim 1, wherein said processing step comprises considering prices in the product information records associated with the products.

9.   A method as recited in claim 1, wherein said processing step comprises considering synonym, hyernym and hyponym relationships in descriptions of the products in the product information records.

10.   A method as recited in claim 1, wherein said processing step comprises considering merchant coverage indicated in the product information records.

11.   A method as recited in claim 1, wherein said processing step includes the steps of grouping said products into subgroups and/or super groups.

12.   A method as recited in claim 1, wherein said determining step comprises:
       scraping attribute values from plural product information records in a group
       and assigning a confidence rating to each scraped attribute value; and
       merging the attribute values into a set of product specification attributes based
on the confidence ratings.

13.   A method as recited in claim 1, further comprising determining a product name for each identified product.

14.   A method as recited in claim 13, wherein said step of determining a name comprises:

   selecting the best name of multiple variant product names from product information records in a group;

   cleansing the best name of superfluous and concatenated text; and

formatting the cleansed name into a product name that is of a predetermined style.

15.   A method as recited in claim 1, further comprising determining a product image for each identified product by selecting a most preferable product image from said product information records.

16.   A method as recited in claim 1, further comprising generating a description of each identified product at least in part from said determined attributes.

17.   A method as recited in claim 1, wherein said processing step comprises examining identification codes associated with each product information record by parsing identification codes present in the product information records and comparing said parsed identification codes to determine commonalities between them.

18.   A method as recited in claim 17, wherein each product information record is examined more than once to determine a common identification code associated with each product.

19.   A method as recited in claim 1, further comprising the step of repeating said processing step after said comparison step, and then performing said comparison step again.

20.   A method as recited in claim 1, further comprising the steps of determining when an outcome of one or more of said processing, correlating, comparing and determining steps falls below a predetermined confidence level and flagging said outcome for further processing.

21.   A method as recited in claim 20, wherein said flagged outcome is deferred and saved and re-processed when more product information sources become available.

22.   A method as recited in claim 20, wherein said flagged outcome is moved to a processing tool for a manual operation.

23.   A method of creating a product catalog stored on computer readable media by aggregating product information from a plurality of product information sources having disparate formats for product information, said method comprising:

    processing plural product information records from the product information sources into one or more groups based on which product information records are likely to correspond to the same product;

    correlating a unique product ID corresponding to an identified product for each of said groups;

    comparing each identified product to categories of in a taxonomy to determine a category for the identified products in the taxonomy;

    repeating the processing and correlating steps after performing the comparing step to revise which groups said plural product information records fall into;

    determining attributes for each categorized product based on the product information records corresponding to each group;

    creating product specifications based on the determined attributes; and

    storing the product specifications in the corresponding determined categories of the taxonomy.

24.   A method as recited in claim 23, wherein said processing step includes determining which products referred to in said product information records are likely to be the same by comparing data strings associated with the products, and determining a common data string.

25.   A method as recited in claim 24, wherein said data strings include at least one of manufacturer part numbers, model identifiers and uniform product codes.

59

26. A method as recited in claim 24, wherein said processing step comprises comparing product names in the product information records and grouping together all products having a substantially similar product name.

27. A method as recited in claim 23, further including the steps of assigning a clustering confidence score to the grouping of information produced by the processing step, and a categorizing confidence score to the categories produced by the comparing step, and repeating said repetition step until said confidence scores stabilize.

28. A method as recited in claim 27, further including the step of flagging the outcome of one or both of the processing step and categorizing step when the confidence score associated with one or both steps falls below a predetermined minimum.

29. A method as recited in claim 28 wherein said flagged outcome is deferred and saved and re-processed when more product information sources become available.

30. A method as recited in claim 28, wherein said flagged outcome is moved to a processing tool for a manual operation.

31. A method as recited in claim 23, wherein said comparing step includes the steps of examining the attributes and attribute value sits for each category, and examining actual product information records already classified in each category.

32. A method as recited in claim 23, wherein the correlating step assigns a different product ID to the same products of different colors.

33. A method as recited in claim 1, wherein said processing step comprises generating a crawler from a server to the product information sources.

34. A method of aggregating product information from a plurality of product information sources in a networked computer environment comprising the steps of:

60

generating a crawler from a server interconnected to the network computer environment to visit the plurality of sources;

gathering product phrase information and characteristics of said product phrase information from each of the plurality of sources via said crawler; and

creating a catalog of products based on the product phrase information and characteristics of said product phrase information.

35. The method of claim 34, wherein said at least one characteristic of said phrase includes at least one of frequency, location, font size, font style, font case, font effects, font color, collocation and co-occurrence of said phrase in each of said plurality of sources.

36. The method of claim 35, wherein the plurality of sources include at least one of a manufacturer's product specifications source, a product literature source, and a merchant's information source.

37. The method of claim 30, wherein said crawler includes a product literature crawler that gathers product phrase information from at least one of said manufacturer's product specifications source and said product literature source.

38. The method of claim 34, further comprising the step of comparing said product phrase information to a categories of a taxonomy to determine product category of products described by the product phase information.

39. The method of claim 1, further comprising determining allied products for at least one of the products.

40. The method as recited in claim 39, wherein said step of determining allied products comprises:

parsing at least one product information record corresponding to a product;

if there is a link in the product information record to related products, following the link to a related product information record;

reverse looking up references to the product in the related product information record; and

relating the related product in the related product information record to the product in the catalog.

41.   The method of claim 23, further comprising determining allied products for at least one of the products.

42.   The method as recited in claim 23, wherein said step of determining allied products comprises:

parsing at least one product information record corresponding to a product;

if there is a link in the product information record to related products, following the link to a related product information record;

reverse looking up references to the product in the related product information record; and

relating the related product in the related product information record to the product in the catalog.

43.   The method of claim 34, further comprising determining allied products for at least one of the products.

44.   The method as recited in claim 34, wherein said step of determining allied products comprises:

parsing at least one product information record corresponding to a product;

if there is a link in the product information record to related products, following the link to a related product information record;

reverse looking up references to the product in the related product information record; and

relating the related product in the related product information record to the product in the catalog.

45.   A system for creating a product catalog by aggregating product information from a plurality of product information sources having disparate formats for product information and storing the information in a taxonomy, said method comprising:

means for processing plural product information records from the product information sources into one or more groups based on which product information records are likely to correspond to the same product;

means for correlating a unique product ID corresponding to the product associated with each of said groups to identify the product;

means for comparing each identified product to categories of a taxonomy to determine a category for the identified products in the taxonomy; and

means for determining attributes for each categorized product based on the product information records corresponding to each group;

means for creating product specifications based on the determined attributes; and

means for storing the product specification in the corresponding determined categories of the taxonomy.

46.   A system as recited in claim 45, wherein said means for processing includes means for determining which products referred to in said product information records are likely to be the same by comparing data strings associated with the products, and means for determining a common data string.

47.   A system as recited in claim 45, wherein said data strings include at least one of manufacturer part numbers, model identifiers and uniform product codes.

48.   A system as recited in claim 45, wherein said means for processing comprises comparing product names in the product information records and means for grouping together all products having a substantially similar product name.

49.   A system as recited in claim 48, wherein said product names are compared without regard for differences in capitalization and punctuation.

63

50. A system as recited in claim 48, further comprising means for parsing generic nouns associated with the products and excluding the generic nouns from processing by said means for determining.

51. A system as recited in claim 45, further comprising means for parsing selected adjectives and excluding the selected adjectives form processing by said determining means.

52. A system as recited in claim 45, wherein said means for processing comprises means for considering prices in the product information records associated with the products.

53. A system as recited in claim 45, wherein said means for processing comprises means for considering synonym, hyernym and hyponym relationships in descriptions of the products in the product information records.

54. A system as recited in claim 45, wherein said means for processing comprises means for considering merchant coverage indicated in the product information records.

55. A system as recited in claim 45, wherein said means for processing comprises means for grouping said products into subgroups and/or super groups.

56. A system as recited in claim 45, wherein said means for determining comprises:

  means for scraping attribute values from plural product information records in a group and assigning a confidence rating to each scraped attribute value; and

  means for merging the attribute values into a set of product specification attributes based on the confidence ratings.

57. A system as recited in claim 44, further comprising means for determining a product name for each identified product.

58.   A system as recited in claim 57, wherein said means for determining a name comprises:

        means for selecting the best name of multiple variant product names from product information records in a group; and

        means for cleansing the best name of superfluous and concatenated text; and formatting the cleansed name into a product name that is of a predetermined style.

59.   A system as recited in claim 45, further comprising means for determining a product image for each identified product by selecting a most preferable product image from said product information records.

60.   A system as recited in claim 45, further comprising means for generating a description of each identified product at least in part from said determined attributes.

61.   A system as recited in claim 45, wherein said means for processing comprises means for examining identification codes associated with each product information record by parsing identification codes present in the product information records and comparing said parsed identification codes to determine commonalities between them.

62.   A system as recited in claim 61, wherein said means for examining examines each product information record more than once to determine a common identification code associated with each product.

63.   A system as recited in claim 45, further comprising means for determining when an outcome of one or more of said processing, correlating, comparing and determining steps falls below a predetermined confidence level and flagging said outcome for further processing.

64.   A system as recited in claim 63, wherein said flagged outcome is deferred and saved and re-processed when more product information sources become available.

65

65.   A system as recited in claim 63, wherein said flagged outcome is moved to a processing tool for a manual operation.

66.   A system for aggregating product information from a plurality of product information sources in a networked computer environment of a comprising:

    means for generating a crawler from a server interconnected to the network computer environment to visit the plurality of sources;

    means for gathering product phrase information and characteristics of said product phrase information from each of the plurality of sources via said crawler; and

    means for creating a catalog of products based on the product phrase information and characteristics of said product phrase information.

67.   The system of claim 66, wherein said at least one characteristic of said phrase includes at least one of frequency, location, font size, font style, font case, font effects, font color, collocation and co-occurrence of said phrase in each of said plurality of sources.

68.   The system of claim 67, wherein the plurality of sources include at least one of a manufacturer's product specifications source, a product literature source, and a merchant's information source.

69.   The system of claim 66, wherein said crawler includes a product literature crawler that gathers product phrase information from at least one of said manufacturer's product specifications source and said product literature source.

70.   The system of claim 66, further comprising the means for comparing said product phrase information to a categories of a taxonomy to determine product category of products described by the product phase information.

71.   The system of claim 45, further comprising means for determining allied products for at least one of the products.

72.   The system as recited in claim 71, wherein said means for determining allied products comprises:

means for parsing at least one product information record corresponding to a product and if there is a link in the product information record to related products, following the link to a related product information record;

means for reverse looking up references to the product in the related product information record; and

means for relating the related product in the related product information record to the product in the catalog.

67